Ownership protection of outsourced biomedical time series data based on optimal watermarking scheme in data mining

Trung Pham Duy
University of Canberra, Australia
duy.pham@canberra.edu.au

Dat Tran
University of Canberra, Australia

Wanli Ma
University of Canberra, Australia

Abstract

In the biomedical and healthcare fields, the ownership protection of the outsourced data is becoming a challenging issue in sharing the data between data owners and data mining experts to extract hidden knowledge and patterns. Watermarking has been proved as a right-protection mechanism that provides detectable evidence for the legal ownership of a shared dataset, without compromising its usability under a wide range of data mining for digital data in different formats such as audio, video, image, relational database, text and software. Time series biomedical data such as Electroencephalography (EEG) or Electrocardiography (ECG) is valuable and costly in healthcare, which need to have owner protection when sharing or transmission in data mining application. However, this issue related to kind of data has only been investigated in little previous research as its characteristics and requirements. This paper proposes an optimized watermarking scheme to protect ownership for biomedical and healthcare systems in data mining. To achieve the highest possible robustness without losing watermark transparency, Particle Swarm Optimization (PSO) technique is used to optimize quantization steps to find a suitable one. Experimental results on EEG data show that the proposed scheme provides good imperceptibility and more robust against various signal processing techniques and common attacks such as noise addition, low-pass filtering, and re-sampling.

Keywords: Ownership protection; Biomedical data; Watermarking; Data mining; Particle Swarm Optimization (PSO).

1 Introduction

In recent years, with continuous development of information technology, an increasing number of hospitals and clinics has digitized patient's medical records and stored these records on independent data servers using e-health technology (Li et al. 2011). Many countries such as Australia, Germany, Taiwan, and the USA have already deployed e-health technology in their national healthcare network. Data sharing or information sharing is necessary for distributed systems in data mining. In healthcare, each collaborator (hospital) needs to share their private local databases with other collaborators in telemedicine system, telediagnosis, telesurgery, and hospital information system (Bhatnagar and Wu 2013). However, there are multiple danger zones like copyright and integrity violations of digital objects (Gupta and Raval 2012), and the sharing of medical data also exposes data holders to threat of data theft (Bertino et al. 2005). In reality, lots of incidents can be found where e-health systems become found wanting due to the miss-management and privacy violations of sensitive health data. (Clearinghouse 2017) provides significant evidence for the above fact and reports more than 44 million healthcare related privacy violations even with the existence of regulations for security and privacy of health information. The Verizon Data Breach Investigation Review (Solutions 2015) states that in 2014 there was only 26 reported incidents of breaches in the healthcare industry. The number of reported data breaches increased 900% by July 2015 to 234 data breaches. USA

health insurer Anthem had a massive breach which resulted in the leak of 80 million client records (Westin 2015). The average annual cost of data breaches to the USA healthcare industry is as high as $7 billion (Smith and Gotel 2007). Moreover, in (Chen and Xu 2013) further statistics reveal the privacy violations by providing examples where personal health information is stolen or acquired without the authorization of legally obliged parties.

Data owners, nonetheless, also need to maintain the principal rights over the datasets that they share, which in many cases have been obtained after expensive and laborious procedures. For example, in teleradiology, one of the most successful e-health services at present, security and privacy protection has become a critical issue (Prior et al. 2009; Ruotsalainen 2010). Medical information security requirements are generally defined by the strict ethics and legislative rules of the security policy/profile, and concerned entitles must adhere to them. Some countries have their own security and privacy policy; for example, Health Insurance Portability and Accountability Act (HIPAA) in USA (Centers for and Medicaid 1996), Code of Federal Regulations number 45 (CFR 45) (Health and Human 2009), and Europe's Directive 95/46/EC (Hänsch and Serna) are expression of such a constraint. In Australia, the Australian Law Reform Commission (Australia. Law Reform 2002) produced the Australian Privacy Law and Practice Report that is a comprehensive review of the Privacy Act of 1988. That reviews incorporates privacy regulations on electronic health information systems.

Cloud computing also significantly contributes to the development of e-health solutions, however considering the fact that cloud infrastructures are managed by third parties who may be curious about the data being stored, confidentiality, integrity and privacy concerns have been raised on the transmitted and stored data. It is necessary to have a right-protection mechanism that can provide detectable evidence for the legal ownership of a shared dataset, without compromising its usability under a wide range of data mining and cloud computing. In such scenarios, the shared data might be illegally sold to third parties by an unauthorized party. In order to cater for such a situation, the data should be right protected so that an unauthorized party might be sued in a court of law. This is only possible, if the data owner is able to prove that the illegally sold data are his property. Therefore, it is important to not only protect the privacy of patient (Alhaqbani and Fidge 2008), but also the ownership (copyright) of the medical data shared with collaborative partners or third party vendors. Therefore, it is important that medical data should be right protected in a manner where ownership could unambiguously be determined (Kamran and Farooq 2012). The important requirement regarding shared medical data is data ownership (copyright) protection. We need effective mechanisms to establish and protect the holders' rightful possession of the data.

Two standard methods, namely encryption and watermarking, are currently used to achieve this aim (Latifpour et al. 2015). In encryption methods, an encryption algorithm is used to encode information, and a receiver at the opposite site recovers the information with a known decoding key. Although the information is encoded on transmission, there will be no guarantee for security of the information after it is decoded at the receiver. For solving such a problem, watermarking technique is introduced to provide the information security after decoding. As an alternative or complement to cryptography, watermarking is mainly used for the copyright protection. In other words, the watermarking is a technique that is able to embed a series of hidden information into an original signal, and to realize the security of information against unauthorized copying, and false claim of owning and exclusiveness. The role of watermarking becomes increasingly important because of the ease of data sharing, particularly through data clouds. Watermarking has many applications such as ownership identification, proof of ownership, tamper detection and leak identification.

1.1 Current issues in sharing and storing biomedical data

One of the major technological and ethical issues governing electronic records is the issue of data privacy. Tampered data can lead to false alarms or incorrect diagnoses of patient. As a result, copyright protection, data authentication and security have become challenging issues due to the illegal modification and distribution. As a data hiding and extraction method, digital watermarking is one of the solutions to address this problem and is used for digital rights

protection, ownership verification and security purposes (Bender et al. 1996). For instance, hospital information system addresses security problems and provides confidentiality, integrity and authentication via watermarking. Watermarking applications for medical purposes have been extensively investigated with reference to their security (Arsalan et al. 2012; Fakhari et al. 2011; Ko et al. 2011).

Although digital watermarking techniques represent a viable solution for the problem of enforcing ownership of medical data (Bertino et al. 2005), preliminary studies in watermarking or information hiding techniques have been developed only for embedding text, images, audio or video in to a host signal. Moreover the techniques developed for these databases do not transpose well to other biomedical time series data modalities such as EEG and ECG for the following reasons: 1) The redundancy in a time series of biomedical signals such as EEG or ECG is less compared with an image or audio, and hence embedding data in to these time series data is much more difficult since the reduced redundancy limits possibilities of hiding data and has not been investigated; 2) Audio signal has slow time-varying feature while biomedical signal is the fast changing-time series; 3) The watermark scheme needs to address a significant challenge related to biomedical data that is insertion of a watermark must not result in changing health and medical data of a patient to a level where a decision maker (or system) can misdiagnose the patient; and 4) In most applications, biomedical time series data such as EEG or ECG is recorded from multiple channels and at relatively high sampling frequencies. Some applications require storage and/or transmission of biomedical data recordings over an extended period of time. As a result, biomedical time series data recordings may lead to a large amount of data.

1.2 Watermarking schemes addressing the current issues

According to these above-mentioned biomedical data characteristics, there are three important issues that watermarking schemes need to address. First, watermarking scheme is required to provide trustworthy evidence for protecting the rightful ownership, while the perceptual difference between the watermarked and the original documents should be unnoticeable to the human observer. Second, good watermarking scheme should satisfy the requirement of robustness and resist distortions due to common attacks. The watermark should be detectable and extractable after data manipulations were applied to the watermarked data. Moreover, the inserted watermark should be imperceptible to intruders and they should not be able to corrupt it by launching malicious attack. Last, watermark extraction is not able to have the original biomedical signal as a large size. These issues motivate us to design an appropriate watermarking scheme for outsourced biomedical data without affecting the perceptual quality of the underlying host signal, having high robust and the original biomedical signal is not required at watermark extraction.

On the other hand, transform domains are proven to be more robust toward different attacks (Mousavi et al. 2014), and singular value decomposition (SVD)-based watermarking is one of the most powerful watermarking schemes in this domain. The robust performance of existing SVD-based watermarking methods is not always better than that of frequency-based methods such as Gaussian filtering and noising (Tsai et al. 2012) for most of attacks (Chang et al. 2005) developed in the spatial domain. A better approach to enhance the robustness of SVD-based methods is to employ this transform along with the frequency transform for biomedical data. It has been reported that among the transform domain methods, Discrete Wavelet Transform (DWT) is more suitable for achieving robust watermarking and imperceptible (Mishra et al. 2014). The present work uses DWT-SVD hybrid transform to carry out watermarking embedding for outsourced biomedical data.

In general, regarding the extraction process, the watermarking methods can be classified in two schemes which are non-blind and blind. In the non-blind watermarking scheme, the extractor can extract watermark data using the original signal. As opposed to that, the blind watermarking methods do not need the original signal for extracting the watermark data. Although the non-blind scheme's complexity is low, they suffer from the following two distinct disadvantages (Gupta and Raval 2012):

i. Security compromise: Attacker may claim the ownership by inserting another watermark in the cover object as non-blind detection does not guarantee unequivocal claims of ownership by the content creator.

ii. Practical application constraints: The presence of original content is required during detection for every watermarking application in non-blind scheme. With a large amount of biomedical data, it is very difficult to ensure the present of original biomedical data in non-blind watermarking scheme.

In order to overcome these disadvantages, blind scheme is suitable for biomedical time series data since the original signal is not required.

1.3 Imperceptibility and robustness issues in watermarking techniques

An acceptable watermarking technique including biomedical data watermarking needs to satisfy two main requirements which are imperceptibility and robustness (Voyatzis and Pitas 1999). Imperceptibility refers to perceptual quality of the data being protected or the quality of biomedical signal should be retained after adding the watermark. Imperceptibility can be evaluated using both objective and subjective measures. The perceptual difference between original biomedical data and watermarked one can be evaluated by Peak Signal Noise Ratio (PSNR). A larger PSNR value indicates that the watermarked biomedical signal more closely resembles its original signal, meaning that watermarked biomedical signal has better imperceptibility. The watermark scheme needs to address a significant challenge related to biomedical data that is insertion of a watermark must not result in changing health and medical data of a patient to a level where a decision maker (or system) can misdiagnose the patient. If a patient is misdiagnosed, it might not only put his life on risk but also result in significantly enhancing the cost of healthcare. This can be guaranteed by a good imperceptibility. According to (Chen et al. 1998), PSNR above 40 dB indicates a good perceptual imperceptibility.

Robustness is ability to extract a watermark from a watermarked biomedical signal after various signal processing attacks. Biomedical signals are not likely to be subject to the same type of malicious attack as downloaded image, audio or video files. However, attacks such as pre-processing signals or downsampling of large data files to allow more efficient data transmission could be an issue. The robustness of the watermark is verified against different attacks such as low pass filtering, addition of Gaussian noise, different sampling rate, and cropping. It is sufficient if the embedded data is robust to simple signal processing techniques necessary for efficient transmission. Normalized correlation (NC) and bit error rate (BER) are metrics to determine the robustness of watermarking scheme.

1.4 Motivations

An effective biomedical data watermarking scheme must satisfy both imperceptibility and robustness goals. If these two requirements are not well achieved, a poor perceptual imperceptibility leads to the fact that a decision maker (or system) can misdiagnose the patient. Meanwhile intruders should be able to corrupt inserted watermark by launching malicious attack with less robust watermarking scheme, resulting in not providing trustworthy evidence for protecting the rightful ownership.

However, there is a trade-off between these goals. For example, increasing the quality of watermarked signal results in a decrease of robustness against attacks. On the design of watermarking scheme, two significant but conflicting requirements, imperceptibility and robustness, should be taken into consideration because the targeted balance between the two requirements. Hence, dealing with the trade-off as an optimization problem may adaptively achieve the specific balance for the intended application. In order to minimize such conflict, we need to find out an optimal balance of the contradictory watermarking requirements, thus improving watermarking performance.

One recently developed and widely used way is to utilize evolutionary and artificial intelligence methods to view the watermarking problems as an optimization problem (Arsalan et al. 2012).

In this manner, the performance of the watermarking scheme can be improved. Artificial intelligence (AI) techniques based on objective function such as tabu-search (Sriyingyong and Attakitmongcol 2006), differential evolution (DE) (Storn and Price 1997), and genetic algorithm (GA) (Kumsawat 2010) have been demonstrated to be very effective in solving conflicting requirements of watermarking. All these techniques intend to work on minimizing the trade-off between the two mandatory objectives of watermarking which are imperceptibility of the embedded watermark and robustness of the embedding scheme.

A proper balance between imperceptibility and robustness also depends upon embedding strength or the strength-of-the-watermark (scaling factor) (Mishra et al. 2014; Run et al. 2012). A large value of scaling factor favors robustness while a small value favors imperceptibility. Hence, scaling factor governs the trade-off between imperceptibility and robustness. Conventional AI techniques have advantages in computing speed for watermarking as they do not bother about scaling factor provided by the user in advance. Yet conventional AI methods cannot handle the automatic balance between imperceptibility and robustness in watermarking. There is no exact algorithm to choose the value of scaling factor. Most of them are based on trial-and-error method, others use the fixed parameters without any optimization in the literature (Wang et al. 2011). Instead of using the fixed parameters, another trend is witnessed, intelligent systems and evolutionary algorithms are integrated into watermarking approach to optimize the parameters. In these artificial intelligence techniques, the meta-heuristic algorithm is chosen to find the suitable scaling factor to overcome the optimization problem. There are several kinds of meta-heuristic algorithms, for example genetic algorithm (GA) (Holland 1992), ant colony (Dorigo et al. 1999), and particle swarm optimization (Kennedy 2011).

Particle swarm optimization (PSO) is another important meta-heuristic algorithm since it has fast convergence and few parameters to adjust (Kennedy et al. 2001). PSO is simple to implement and computationally efficient as well. Basic principle of the PSO algorithm is the clever exchange of information between the global and local optimal values. A global optimum is achieved by updating generations based on movement and intelligence in an evolutionary system. In a similar pattern of other evolutionary and stochastic computation techniques, such as GA and tabu-search, PSO shares similar characteristics, but it was proven that PSO could obtain better results in a faster and cheaper way compared with GA methods. Besides, another fascinating feature of PSO is that there are fewer parameters to adjust compared with GA methods (Lei et al. 2013). Due to its advantage over other global optimization techniques, a myriad of watermarking scheme is developed and implemented based on PSO in this research.

1.5 Research focus and contribution

Our current research work focuses on optimizing the trade-off between the two problems in biomedical data watermarking: imperceptibility/visual quality of the signed or attacked biomedical signal and the issue of robustness. We propose a blind biomedical data watermarking scheme based on the hybrid DWT-SVD transform and identify the optimal quantization using PSO. Recent studies prove that computational intelligence techniques, especially PSO (Hassan et al. 2005) are ideally suited for solving constrained optimization problem in real-time. PSO has the following advantages: 1) PSO is easier to implement and there are fewer parameters to adjust; 2) PSO has a more effective memory capability since every particle remembers its own previous best value as well as the neighborhood best; and 3) Since all the particles use the information related to the most successful particle, PSO is more efficient in maintaining the diversity of the swarm. In order to achieve a trade-off between imperceptibility and robustness, PSO is utilized to search for the optimal embedding parameter in our study. An objective function used in this algorithm is a linear combination of PSNR and NC. The NC is a metric to determine the robustness and therefore the research selects three different signal processing operations (additive noise, re-sampling, and low-pass filtering) as attacks to evaluate it. Experimental results show that our proposed watermarking scheme yields a good imperceptibility and more robust against various signal processing and common attacks.

The main contributions of this research are as follows

1. Developing an appropriate efficient blind watermarking algorithm that is suitable for outsourced time series biomedical data such as EEG or ECG in data mining; and
2. Determining the optimal balance of the contradictory watermarking requirements using PSO to find out the optimal quantization steps for different host time series biomedical signals and watermarks.

In addition, chaotic map with the chaotic characteristic is also used to enhance the confidentiality of the proposed watermarking scheme. With the proposed blind watermark detection algorithm, the biomedical watermarking scheme can extract the watermark thus saving a lot of space for storing the original biomedical data and watermark.

The rest of this paper is organized as follows. In Section 2, we review the state-of-the-art biomedical data watermarking schemes. Section 3 introduces the related background of our proposed scheme including chaotic encryption, singular value decomposition (SVD), discrete wavelet transform (DWT), and particle swarm optimization algorithm (PSO). Section 4 presents details of the proposed biomedical data watermarking scheme. Our experimental results are presented and analyzed in Section 5. Finally, Section 6 draws conclusions and presents discussions on the future work.

2 Related Work

Biomedical data sharing via distributed systems and data clouds in healthcare introduces new security and privacy threats as well as data integration issues. HIPAA mandates that while transmitting information, a patient's privacy and confidentiality must be protected (Lee and Lee 2008). In addition, it is of crucial importance to implement a security protocol which will have powerful communication and storage security (Malasri and Wang 2007). Encryption and cryptographic algorithms have been used to secure data during the communication and storage. As a result, the final data will be stored in encrypted format (Hu et al. 2007; Maglogiannis et al. 2009; Wang et al. 2010). Cryptography is the most common method of protecting digital content. However, once an encrypted data segment is decrypted, its content has no protection. In addition, large computational overhead is needed in encryption-based techniques. As an alternative or complement to cryptography, data can be hidden in other innocuous-looking objects so that their presence is not revealed through steganography or watermarking. Watermarking techniques have been thoroughly studied as a means to achieve proof of ownership and transaction tracking (Cox et al. 2007). Encryption can offer confidentiality and integrity in content protection, and the decrypted content can be further protected using digital watermarking. Watermarking techniques for embedding text, images, audio or video in a host signal have been developed. However, there has been little research on watermarking for time series biomedical data. Based on the embedding information concept, watermarking algorithms can be broadly classified as either spatial or transform domain (Zain and Clarke 2011). Spatial domain algorithms directly insert the watermark into the biomedical signal, whereas transform domain algorithms embed the watermark based on a modified version of the biomedical signal. In spatial domain, (Kong and Feng 2001) proposed an elementary watermarking technique for ECG signals. They describe three popular watermarking techniques applied to EEG signals: Benders Patchwork (Bender et al. 1996), Least Significant Bit (LSB) (Van Schyndel et al. 1994), and Quantization Index Modulation (QIM) (Chen and Wornell 2001) watermarking with regard to their ability to verify EEG signal integrity after noise contamination resulting from communication. LSB watermarking scheme in support of proof-of-ownership for ECG signals is proposed in (Ibaida et al. 2011). However, LSB watermarks provide poor robustness to malicious alterations. The authors of (Kaur et al. 2010) describe a spread spectrum watermarking scheme that embeds robust and imperceptible watermarks into ECG signals. However, such a scheme addresses security considerations only during communication of the data rather than over the course of sharing it. These spatial domain approaches are fast, simple and provide high capacity for embedding watermarks, however, they cannot survive or less robust against noise or signal processing

attacks (Zain and Clarke 2011). Furthermore, once the method is uncovered, embedded watermark can be easily modified by a third party. These techniques are simple and straightforward but do not strongly ensure protection against removal and robustness of the watermarked biomedical data.

Due to the shortcomings of watermarking in the spatial domain, most watermarking techniques operate on a transformed domain to provide high robustness. The robustness can be increased by making the human visual system (HVS) less sensitive (Heylen and Dams 2008). To achieve this, prior to embedding the watermark, transformation such as discrete Fourier transform (DFT), discrete cosine transform (DCT), DWT, or SVD is applied onto host signal to obtain the watermarked signal. In addition, the watermark embedded in the transform domain is more imperceptible than that in time domain according to theoretical analysis and simulation results (Chen et al. 2014). (Engin et al. 2005) proposed a wavelet transformation based watermarking technique for ECG. They used techniques like scrambling matrix to shuffle the secret data information embedded in the cover signal. (He et al. 2012) proposed a self-synchronized watermark technology to protect the ECG. Their study confirmed that the use of wavelet-based quantization watermarking on ECG signal is adequate for patient protection. (Ramu et al. 2016) proposes ECG stenography using Continuous Ant Colony Optimization and DWT-SVD watermark embedding techniques. The authors of (Pham et al. 2015) proposed an approach that uses discrete wavelet transform (DWT) to decompose EEG signals and SVD to embed watermark into the decomposed EEG signal. (Jero et al. 2014) proposed a Steganography approach based on DWT and SVD to hide patient confidential information along with the ECG data. Generally, these above-mentioned watermarking methods usually employ pre-defined embedding rules to determine their embedding parameters, such as embedding strengths, and thresholds, either empirically or experimentally. Since the embedded results relying on the predefined rules always show the same performance, these watermarking schemes cannot approach the inherent performance upper limit. If the performance of embedded outcome is not satisfactory in some respects, the only solution is to adjust the embedding parameters empirically so that better watermarked results are obtained. However, empirically settings the parameters is an awkward process that lacks systematic techniques (Huang and Wu 2009). In addition, watermarking algorithms need larger parameter space, therefore, it is often difficult to determine optimal watermarking parameters empirically or experimentally. As a result, these watermarking methods do not exhibit desirable performance (Lai 2011). It is necessary to build a watermarking scheme which aims to provide a robust and adaptive system for biomedical data protection against copyright infringement issues. To attain or approach the upper performance limit of previously watermarking algorithms, we must determine their optimal watermarking parameters. However, as stated above, it is usually difficult to empirically determine optimal watermarking parameters. A popular way of solving the optimal watermarking problem is to consider it as an optimization problem (Wang et al. 2011).

3 Background

3.1 Chaotic encryption

Chaotic maps have been frequently used in digital watermarking. Chaotic encryption of a watermark image is performed using Arnold transform (Zheng et al. 2007) which is also called Cat Face transfer and is given by

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} (mod\ N) \quad (1)$$

where (x, y) is the pixel of the watermark image, (x', y') is the pixel of the watermark image after scrambling, and N is order of watermark image matrix. Since the Arnold transform is periodic, the scrambling number can be considered as the key to enhance the security.

3.2 Singular Value Decomposition (SVD)

Let $A = (a_{ij})_{N\times N}$ be an $N\times N$ matrix. The SVD of matrix A is represented in the form $A = USV^T$, where U and V are orthogonal matrices, S is a diagonal matrix with nonnegative elements, and superscript T denotes matrix transposition.

The diagonal elements of S, denoted by $\sigma_i, 0 \le i < N$, are called the singular values (SVs) of A and are assumed to be arranged in decreasing order, i.e., $\sigma_i > \sigma_{i+1}$. The columns of U, denoted by $U_i, 0 \le i < N$, are called the left singular vectors, while the columns of V, denoted by V_i, are called the right singular vectors of A. The SVD has some interesting properties as follows: (i) the sizes of the matrices for SVD transformation are not fixed, and the matrices need not to be square, (ii) changing SVs slightly does not affect the quality of the signal much, (iii) the SVs are invariant under common signal processing operations, and (iv) the SVs satisfy intrinsic algebraic properties.

3.3 Discrete Wavelet Transform (DWT)

DWT employs extensive time window for low frequencies and short time window for higher frequencies. DWT is widely used for the time-frequency analysis of biomedical signals (Jahankhani et al. 2006; Orhan et al. 2011), especially in a time series biomedical data such as EEG signal analysis due to its non-stationary characteristics. As EEG signal is the fast changing-time series with continuously random changes, we use the Haar wavelet which is more suitable for such fast changing time-series compared to Daubechies wavelets, Mexican Hat wavelets and Morlet wavelets which are better suited for smoothly changing time series (Percival and Walden 2006). In addition, the Haar wavelet is also simple, fast and exactly reversible which is necessary to reconstruct cover signal in digital watermarking. Each wavelet decomposition of the original signal halves the frequency and length of the signal. The Haar function Ψ used as the mother wavelet generates a set of wavelets as follows:

$$C_{a,b} = \sum_{N_{samp}} c(t)\Psi_{a,b}(t) \quad (2)$$

Where $\Psi_{a,b}(t) = \frac{1}{\sqrt{s}}\Psi_{a,b}\left(\frac{t-\tau}{s}\right)$, a denotes the dilation index, b the translation index, s the scale factor and τ the displacement.

3.4 Particle Swarm Optimization Algorithm (PSO)

PSO (Kennedy 2011) is a population-based stochastic algorithm developed for continuous optimization. In PSO, each particle which represents a potential solution will search for optimal coordinates in the problem space. Each particle has its own set of attributes including *position, velocity,* and a *fitness value* which is obtained by evaluating a fitness function at its current position. The algorithm starts with the initialization of particles with random position and velocities so that they can move in the solution space. Then, these particles search the solution space for finding better solutions. Each particle keeps track of its *personal best position* found so far by storing the coordinates in the solution space. The best position found so far by any particle during any stage of the algorithm is also stored and is termed as the *global best position*. The velocity of every particle is influenced by its personal best position (autobiographical memory) and the global best position (publicized knowledge). The new position for every particle is calculated by adding its new velocity value to every component of its position vector.

Let D denote the swarm size. Each individual particle $i, 1 \le i \le D$ has its own position p_i and velocity v_i. These particles search for the optimal value of a given objective function iteratively, then locate their individual best positions p_i^{best} (*pbest*) and keep track of the global position $p_{g_i}^{best}$ (*gbest*) from all best positions through a search space. With respect to the two best values, the velocity and position of particle $i(1 \le i \le D)$ in iteration t+1 are updated by:

$$p_i(t+1) = p_i(t) + v_i(t+1) \quad (3)$$

$$v_i(t+1) = c_0 v_i(t) + c_1 r_1 \left(p_i^{best}(t) - x_i(t)\right) + c_2 r_2 \left(p_{g_i}^{best}(t) - x_i(t)\right) \quad (4)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (5)$$

where c_1 and c_2 are referred to as cognitive and social parameters, respectively, which are positive constants, r_1 and r_2 are random numbers uniformly distributed in the range of [0, 1]. c_0 is the inertia weight in the range of [0, 1], which controls the momentum of the particle and tunes changes in these values. These coefficients control how far a particle moves in a single iteration. v_i is the moving distance in one-step for a particle i and is limited to the range of $[v_{min}, v_{max}]$, where v_{min} and v_{max} are the minimum and maximum moving distance in one step, respectively. The qualitative measure of the selection of PSO algorithm parameter can be found in (Trelea 2003). From the PSO equations, it is known that the trade-off is related to the PSO parameters such as c_0, *itermax*, c_1 and c_2. Based on the analysis in (Trelea 2003), the parameter couples that are close to the center of the stability triangle lead to quick convergence, while parameter couples that are close to its borders need many iterations to converge. After going through the whole process iteratively, the evaluated objective function reaches the desired termination criterion.

4 Performance Evaluation Framework

The performance of the proposed watermarking method is mainly investigated by measuring its imperceptibility and robustness. The imperceptibility requirements guarantee that the watermarked biomedical data are useable for diagnosis and other clinical uses. Robustness, basically, is defined as the degree of resistance of a watermark scheme to modifications of host signal due to either common signal processing, or operation devised specifically in order to render the watermark undetectable (Tefas et al. 2009).

4.1 Perceptual transparency/Imperceptibility

The watermark must be embedded without affecting the perceptual quality of the underlying host signal. The procedure is imperceptible if the Human Audio/Visual System (HA/VS) cannot differentiate between the original host signal and a host signal with inserted data. These perceptibility measurement terms are explained as applied to images, but they can also be used with time series data such as EEG and ECG, as the clinician views the ECG and EEG to make a diagnosis. Similar to digital image watermarking, biomedical data such as EEG or ECG is based on human visual system (HVS) and is typically analyzed in two ways: 1) Visual inspection by human experts and 2) automatic analysis using processing algorithms. Watermarking techniques need to reconstruct biomedical data without introducing any errors in such analyses. According to (Planitz and Maeder 2005), biomedical data is mainly used for diagnosis, thus the imperceptibility of the watermark should be as high as possible. Distortions to the original due to the watermark may result in wrong interpretation of the data.

Common trend for analyzing imperceptibility is through objective and subjective methods, and is utilized in the present study as well. As reported in the literature, imperceptibility is through an objective process carried out by considering PSNR in Eq. (6) which evaluates the perceptual difference between original biomedical data and watermarked one.

PSNR(X,X') is used to represent the imperceptibility, which denotes the Peak Signal To Noise Ratio between the original signal X and watermarked signal X'. PSNR(X,X') measure is defined as follows:

$$PSNR = \frac{20 \log_{10} max(x)}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(x - x')^2}} \quad (6)$$

A larger PSNR value indicates that the watermarked biomedical signal more closely resembles its original signal, meaning that watermarked biomedical signal has better imperceptibility. In biomedical watermarking application, perceptual similarity must be very high to avoid any risk of misdiagnosis. Therefore, minimum PSNR of 40–50 dB is advised (Chen and Ramabadran

1994). If the value of PSNR is large, it indicates that the noise or distortion due to the embedded watermark is very small.

4.2 Robustness

Robustness is the resistance of watermark signal against common signal processing and malicious attacks. Medical signals are not likely to be subject to the same type of malicious attack as downloaded image, audio or video files. However, attacks such as pre-processing signals, or downsampling of large data files to allow more efficient data transmission could be an issue. The robustness of the watermark is verified against different attacks such as low pass filtering, addition of Gaussian noise, different sampling rate, and cropping. It is sufficient if the embedded data is robust to simple signal processing techniques necessary for efficient transmission.

Normalized correlation (NC) and bit error rate (BER) are adopted to evaluate robustness in the proposed watermarking scheme. NC between original watermark X and extracted watermark after attack X' is a metric to determine the robustness, and is calculated as follows:

$$NC(X, X') = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} x(n,m) \times x'(n,m)}{\sqrt{\sum_{n=1}^{N}\sum_{m=1}^{M} w(n,m)^2} \times \sqrt{\sum_{n=1}^{N}\sum_{m=1}^{M} w'(n,m)^2}} \quad (7)$$

In Eq. (7), NC is between 0 and 1. If NC is close to 1, the similarity between original watermark X and extracted watermark X' is very high. Otherwise, NC is close to zero, the similarity between X and X' is very low.

The error in extracted watermark bits due to watermarking process can be measured using BER (Loukhaoukha et al. 2011) as given in Eq. (8). Here, W_{ret} is the number of retrieved watermark bits without error; W_{org} is the total number of original watermark bits.

$$BER\ (W, W') = \sum \frac{W_{ret}}{W_{org}} \times 100 = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} w(n,m) \oplus w'(n,m)}{N \times M} \quad (8)$$

where the symbol $\oplus$ is exclusive-or (*XOR*) operator.

The BER measures the difference between an original watermark and the corresponding extracted watermark. A lower BER suggests that the extracted watermark resembles the original watermark more closely. BER is used to evaluate our watermarking methods against signal distortions.

5 The proposed biomedical data watermarking scheme

Our proposed scheme adopts a blind digital watermarking scheme based on DWT-SVD with PSO optimization for time series biomedical data. The proposed method can extract the embedded watermark without any information from the original watermark. The proposed approach includes three primary phases which are watermarking embedding, watermarking extraction and quantization step optimization with PSO. Embedding and extraction phase are presented in Fig. 1 and Fig.2, respectively, and the quantization step optimization with PSO is shown in Fig. 3.

5.1 Watermarking embedding

Watermarking scrambling algorithm is first used in order to dispel the pixel space relationship of the binary watermarking image and to improve the whole digital watermarking system. We use Arnold transform to encrypt the watermark image. In this research, the key for Arnold transform is denoted as *K* which is the number of scrambling.

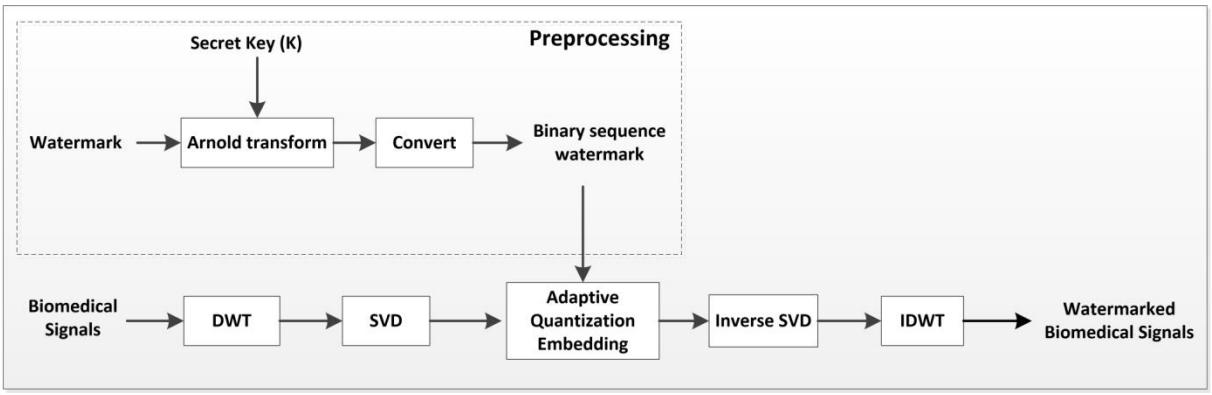

Figure 1. Diagram of embedding the biomedical signal watermark

The steps of the biomedical watermarking embedding are summarized as follows:

Step 1: The watermark W after Arnold transformation is converted into a one-dimensional watermark sequence of length L.

Step 2: The original biomedical signal X is decomposed by two-level DWT using a Haar wavelet filter in order to get three sets of coefficients *D1, D2* and *A2* (detailed and approximate coefficients, respectively). In our observations, larger decomposition level will not increase the watermarking robustness while it causes intensive computation. Thus we choose two as decomposition level of the Haar wavelet to trade-off between robustness and imperceptibility.

Step 3: The set of coefficients *A2*, corresponding to low frequency part, is divided into K non-overlapping segments, so that each watermark bit is inserted into one segment. This ensures a better distribution of watermark bits over the entire biomedical signal, improving the robustness of the watermark against different attacks.

Step 4: Each segment is rearranged into a $(r \times r)$ matrix block named M_l.

Step 5: SVD is performed to decompose each matrix M_l into three matrices: U_l, S_l, and V_l. The SVD operation is represented as follows:

$$M_l = U_l \times S_l \times V_l^T \qquad (9)$$

Step 6: Insert the watermark. Due to the stability of the matrix S_l under different attacks, the insertion of the watermark is performed by manipulating the coefficient in the highest singular value $S_i(1,1)$ of each matrix S_i by adaptive dither modulation (DM) quantization methods. As the first singular values have the highest energy values, they are used to embed the watermark in order to guarantee the robustness and transparency. In addition, the popular DM quantization method has good robustness and blind nature, thus it is used in the embedding process. This embedding strategy can be formulated in the following quantization function:

$$\begin{cases} S'_i(1,1) = round\left[\dfrac{S_i(1,1)}{\Delta}\right] + \dfrac{3}{4}\Delta & if\ w_l = 1 \\ S'_i(1,1) = round\left[\dfrac{S_i(1,1)}{\Delta}\right] + \dfrac{1}{4}\Delta & if\ w_l = 0 \end{cases} \qquad (10)$$

where Δ is quantization step, $round$[.] is rounding to the nearest integer value. It is obvious that the quantization step is significant in terms of both robustness and imperceptibility. The larger the quantization step is, the more robust, but less transparent, the watermarking scheme is. Therefore, the quantization step should be specially developed to achieve optimal performance.

Step 7: Each modified singular value is reinserted into matrix S_i and inverse SVD transformation is conducted to obtain the watermarked block M'_l which is given by:

$$M'_l = U_l \times S'_l \times V_l^T \quad (11)$$

Step 8: The modified blocks are rearranged to one-dimensional vector and concatenated in order to obtain the modified approximation vector A'_2

Step 9: The watermarked signal X' is obtained by calculating the two-level inverse DWT using modified approximation vectors A'_2 and the original detailed vectors D_1 and D_2.

5.2 Watermark extraction

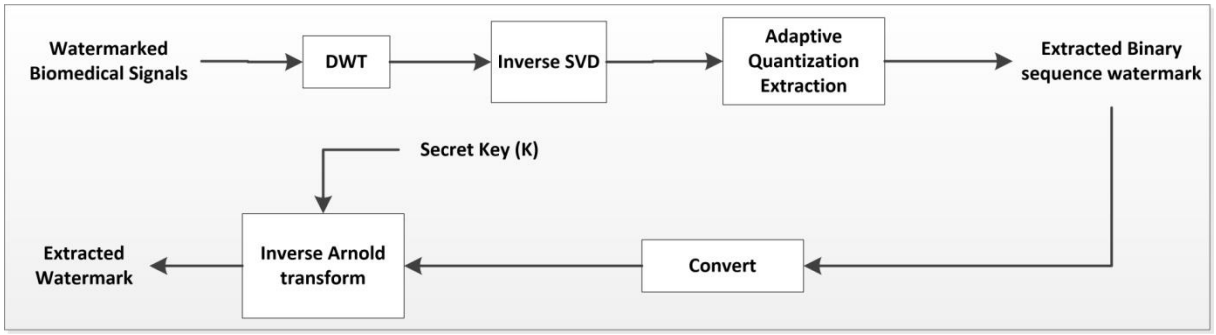

Figure 2. Diagram of the biomedical signal watermark extraction

The detection is rather simple when only the watermarked signal and the watermark embedded positions used in watermark embedding are needed to extract the watermark. The watermarking extraction includes the following steps:

Step 1: Performing two-level DWT to the watermarked and possibly distorted host biomedical signal X' using a Haar wavelet filter to obtain three sets of coefficients D'_1, D'_2 and A'_2.

Step 2: DWT coefficient, A'_2 is divided into different non-overlapping blocks M'_l with the same block length as that in the watermark embedding process.

Step 3: SVD transformation is applied on each block to produce singular values

$$M'_l = U'_l \times S'_l \times V'^T_l \quad (12)$$

Step 4: The largest singular value of each diagonal matrix S'_i located at the same position in the pre-embedding process is calculated.

Step 5: Let $\emptyset = \left(S'_l - floor\left\lfloor\frac{S'_l}{\Delta}\right\rfloor\right) \times \Delta$, where *floor*[.] is the rounding function which rounds the elements to their nearest integer. The embedded meaningful watermark sequence is extracted as follows:

$$\begin{cases} w'_l = 1 \; if \; \emptyset \geq \frac{\Delta}{2} \\ w'_l = 0 \; if \; \emptyset < \frac{\Delta}{2} \end{cases} \quad (13)$$

Step 6: Organizing a 2-D matrix from the watermark sequence, then the binary watermark image W' is obtained by Arnold transformation with key K. The biomedical watermarking extraction phrase does not require original cover signal at the receiver, thus the proposed approach constitutes a blind watermarking scheme.

5.3 PSO Optimization Performance

In the design of biomedical watermarking system, there are two goals that are always conflicted. These goals are imperceptibility and robustness. In order to minimize such conflict, this work employs the PSO algorithm to search for optimal quantization step, thus allowing the system to achieve optimal performance.

In biomedical watermarking, the quantization step is signal-dependent where different biomedical signals require different quantization steps, rather than a fixed one. As seen in insertion process, the watermark is carried into the biomedical signal by quantization of the selected coefficient of each matrix S_l. Large quantization step (high Δ) of coefficients results in better robustness and more distortion of biomedical signal, while a small quantization step (low Δ) leads to low robustness and low distortion. Hence, the parameter Δ must be carefully selected for each biomedical signal, in order to ensure the best performance in terms of imperceptibility and robustness. However, the empirical selection of Δ is not an optimal solution. The selection process must be automatically performed in order to improve the performance of the proposed watermarking method.

It is clear from this discussion that any objective function used to optimize watermark embedding should take both PSNR and NC into account. To solve this problem, the PSO optimization is used to find the adequate quantization step, for each signal, which guarantees the best imperceptibility-robustness compromise. The optimization process for finding the suitable quantization step, Δ in our watermarking scheme is shown in Fig. 3.

The most critical step in the optimization process is the definition of a reliable objective function. An objective function is a fitness measure on solution represented by each chromosome. Its value tells how well the chromosome satisfies the final goal. The objective value function should be designed as a function of both imperceptibility and robustness to obtain the optimal performance by PSO, the objective function can be designed as

$$Objective = f(imperceptibility, robustness) \quad (14)$$

PSNR and NC are employed to represent the imperceptibility and robustness of our scheme. Therefore, the objective function is of the following form:

$$Objective = \max(PSNR + \gamma \times \frac{1}{R}\sum_{i=1}^{k} NC_i) \quad (15)$$

where R is the number of attacks. Weighting factors are introduced as significant difference that might take place between the metrics of the watermarked biomedical host and the extracted watermark. As the PSNR is much larger as compared to the associated NC values therefore a weighting factor γ is used to balance out the influences caused by the two parameters. Generally speaking, the PSNR value should be greater than 40dB (Chen et al. 1998) which guarantees the good perceptual transparency, while the NC in Eq. (7) value lies between 0 and 1. Therefore, there is a strong need to include weighting factor.

The values of quantization step are obtained by implementing PSO algorithm. For each iteration in PSO, the value of Δ is examined for several attacks, such as noise attack, re-sampling attack, and cropping attack. Due to the flexibility of the developed system, the other attacking schemes can easily be added to the system or replaced with those used in the PSO optimization process. At the end of PSO iteration, we will obtain the near optimum quantization step. To find an optimal solution in the objective function, PSO is first initialized with a group of random particles, each of which represents a candidate solution to the problem, then searches for optima by updating generations. In order to gain the optimal performance, f_{obj} should be optimized at PSO processes. Using the above objective function f_{obj}, the quantization step Δ can be optimally searched to achieve the best of both biomedical data signal quality and watermark robustness.

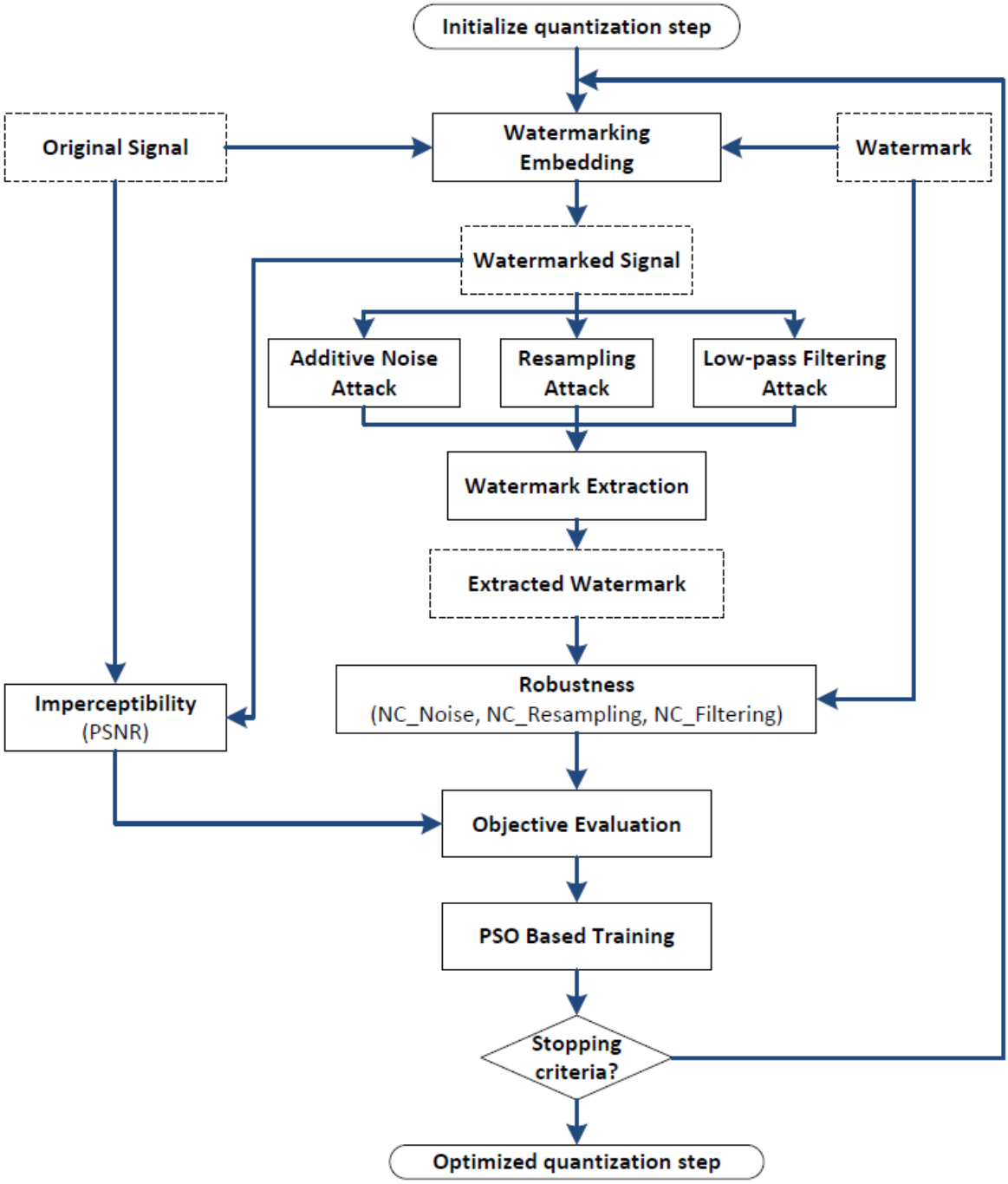

Figure 3. PSO Optimization Performance Process

Fig.3 illustrates the PSO Optimization Performance Process. The main steps can be summarized as follows:

Step 1 (Initialization): Generate an initial population of individuals (quantization steps) and evaluate the fitness values.

Repeat until maximum number of generations reached.

Step 2 (Watermarking Embedding): embed the watermark in cover biomedical signal using the solution (trial individual) (Section 5.1).

Step 3 (Attacks): applying the attacks (additive noise, re-sampling, and low-pass filtering) on watermarked biomedical signal.

Step 4 (Extraction): extract the watermarks from the corrupted watermarked biomedical signals (Section 5.2).

Step 5 (Objective Evaluation): evaluate the robustness between the watermark and extracted ones, and imperceptibility between watermarked signal and original one. Using these compute the fitness value/objective , Eq. (15).

Step 6 (PSO Based Training): select the solution for next generation depending on the quality of the solution in PSO training.

Termination: Check the stopping criteria. If yes then stop, the best quantization step (Δ) is found; otherwise go to Step 2.

6 Experiments and Results

In our experiments, the DEAP dataset (Dataset for Emotion Analysis using Electroencephalogram, Physiological and Video Signals) which is an open database proposed by Koelstra et al. (Koelstra et al. 2012) was used as the original EEG signals. The 5 random channels (FP1, FC1, F3, P7 and T7) of 32 subjects were chosen to test. A binary logo image with size 32x32 will be used as the watermark image in Fig. 7(a). We should examine the effect of quantization steps on imperceptibility and robustness to find out the searching range of quantization steps.

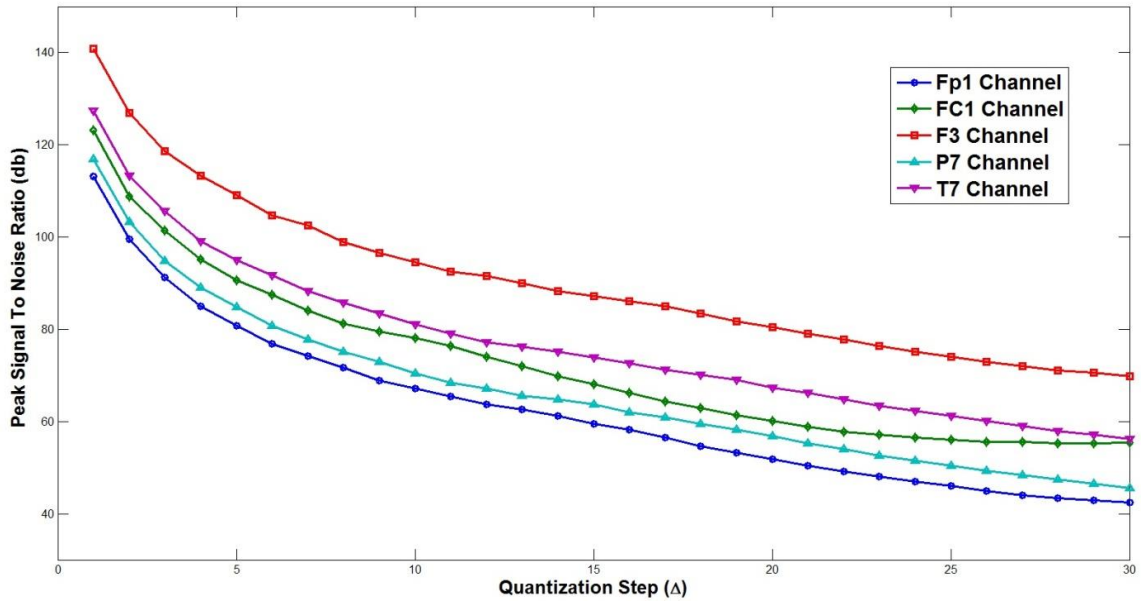

Figure 4. Effect of quantization steps on PSNR in different EEG channels.

To examine the effect of quantization on visual quality of watermarked EEG signals, we plot its corresponding computed values as a function of Δ. Figure 4 depicts the effect of quantization steps on *PSNR* for 5 random EEG channels. It is clear that *PSNR* and Δ are inversely proportional to each other for all channels. (Chen and Ramabadran 1994) suggest the minimum PSNR of 40–50 db in their research, hence, quantization steps in Figure 4 should be less than 30.

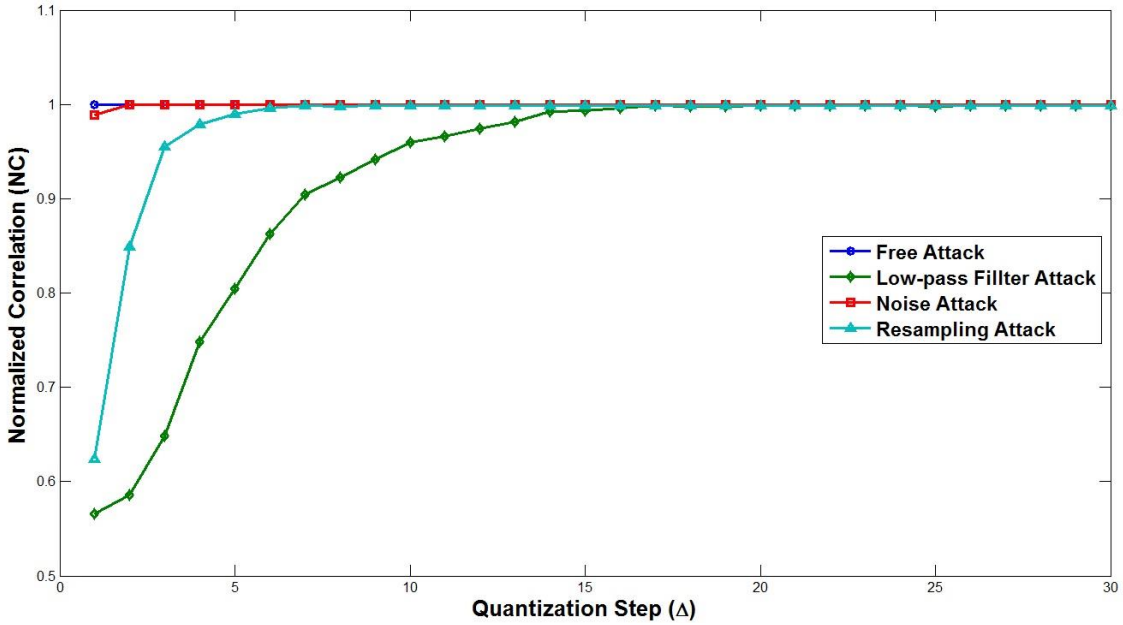

Figure 5. Effect of quantization steps on NC values with different attacks

We investigated the variation of NC(X, X') with respect to quantization steps (Δ). Figure 5 plots the effect of quantization steps on NC(X, X') value of Fp1 channel of Subject 1 with different attacks. It is clear from this figure that NC(X, X') values vary only between the range of Δ which is [1, 20]. With $\Delta > 20$, the NC(W, W) values get stabilized. This is specifically true for all attacks including noise addition, re-sampling and low-pass filtering. For these attacks, the NC(X, X') values with respect to Δ are throughout constant. Thus, it is concluded that for the attacks used by us in this simulation, the range [1, 20] of Δ is an appropriate range to determine the suitable quantization step. We therefore use this range for Δ for our future computations for Fp1 channel of Subject 1.

Based on empirical experience and the trial and error, the PSO optimization parameters were chosen to achieve the optimal robustness and transparency as follows $c_0 = 0.4$, $c_1 = c_2 = 1.8$, number of particles = 30, number of generation = 50, and weighting factor $\gamma = 50$.

6.1 Imperceptibility

In our scheme, *PSNR* was employed to evaluate the differences between original EEG signals and watermarked EEG signals. It should be noted that the larger *PSNR*, the better imperceptibility. A larger *PSNR* value indicates that the watermarked EEG signal more closely resembles its original signal, meaning that watermarked EEG signal has better imperceptibility.

EEG Channel	Quantization step (Δ)	PSNR (in dB)	NC	BER
FP1	19.867	60.558	1	0
FC1	19.476	64.829	1	0
F3	19.573	59.118	1	0
P7	20.925	66.395	1	0
T7	19.978	70.275	1	0
Average	**19.964**	**64.235**	**1**	**0**

Table 1. Performance metrics on average for different EEG signal channels of 32 subjects

According to (Chen et al. 1998), PSNR above 40 dB indicates a good perceptual fidelity. The PSNR values (in dB) of the watermarked EEG are shown in Table 1, all of them are higher than 40 dB, thus this indicates that diagnosability is not lost and degradation to the overall signal is acceptable. The comparable results in terms of PSNR and Bit Error Rate (BER) are listed in Table 2. The results clearly show that the proposed method outperforms the other three methods.

	Characteristics	PSNR (in dB)	BER
Proposed scheme	***Scheme:*** *DWT-SVD with PSO* ***Data Type:*** *EEG signal*	**64.235**	**0**
Pham's method (Pham et al. 2015)	***Scheme****: DWT-SVD* ***Data Type****: EEG signal*	57.53	N/A
Ramu's method (Ramu et al. 2016)	***Scheme****: DWT-SVD with Continuous Ant Colony Optimization (CACO)* ***Data Type****: ECG signal*	62.87	0
Jero's method (Jero et al. 2014)	***Scheme:*** *DWT-SVD* ***Data Type:*** *ECG signal*	50.44	0

Table 2. The performance comparison results for PNSR and BER

Figure 6 shows the difference between the original EEG signal and watermarked one. It can be seen that this difference is close to zero, it means the distortion to the original EEG due to the watermark is minimal. Therefore, it shows that our watermarked EEG signal is near identical to the original EEG signal.

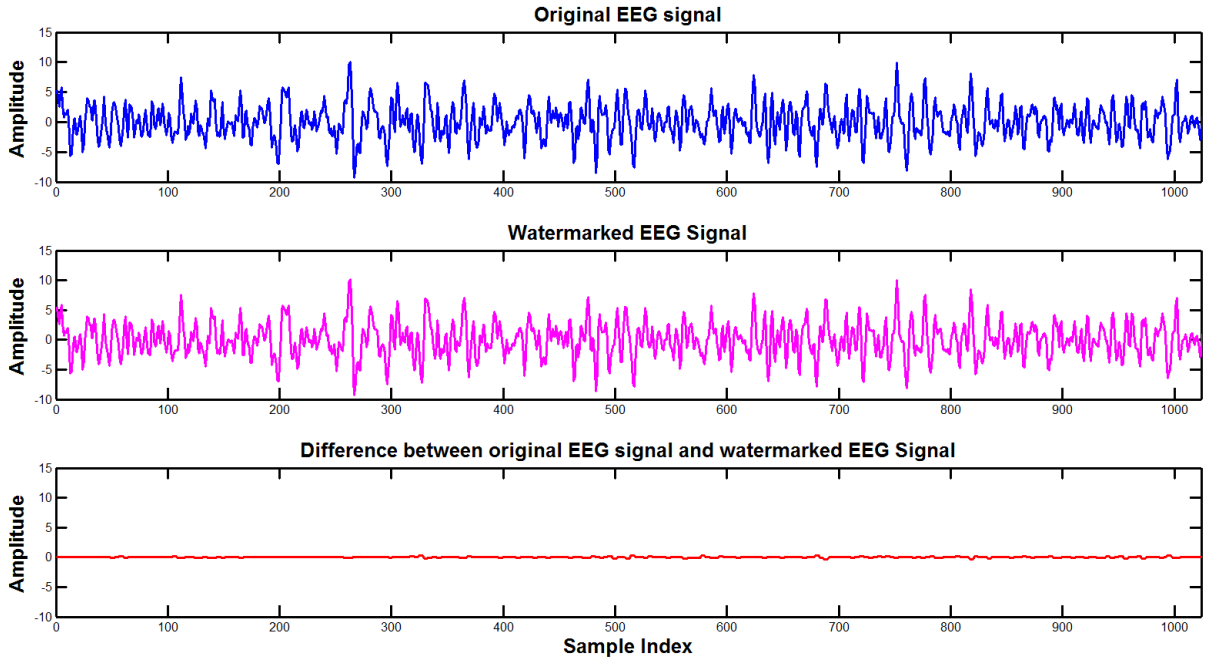

Figure 6: Original EEG signal vs Watermarked EEG signal (above), Difference between original EEG signal and watermarked EEG signal (below) in Channel Fp1 of Subject 01

6.2 Robustness

In order to evaluate the robustness of the proposed method against the common signal processing attacks, we used BER and NC measures.

The following signal attacks were performed in Matlab:

1. Noise addition: Additive white Gaussian noise (AWGN) was added to the watermarked EEG signal with 20dB.
2. Low-pass filtering: The low-pass filter with cut-off frequency of 40Hz was applied to all watermarked EEG signals.
3. Re-sampling: The original EEG signals were sampled with a sampling rate of 128 Hz. Watermarked EEG signals were re-sampled at 64 Hz and then restored by sampling again at 128 Hz.

In attack-free case, we extracted watermark from watermarked EEG signals using the proposed watermark extraction algorithm. Table 1 shows BER = 0, NC = 1 in case of no attack, meaning that watermark can be accurately extracted from the watermarked EEG signal. The signal attacks were performed in Matlab including noise addition, random cropping, low-pass filtering and re-sampling.

EEG Channel	Noise Addition		Low-pass Filtering		Re-sampling	
	BER(%)	NC	BER(%)	NC	BER(%)	NC
FP1	0.32	0.9971	2.93	0.9762	3.71	0.9703
FC1	0.67	0.9947	1.41	0.9892	3.52	0.9715
F3	1.21	0.9861	3.51	0.9718	2.48	0.9802
P7	0.49	0.9961	3.42	0.9726	0.65	0.9947
T7	0.88	0.9929	2.72	0.9812	1.39	0.9869
Average	**0.71**	**0.9933**	**2.80**	**0.9782**	**2. 35**	**0.9807**

Table 3. Performance metrics for different EEG signal channels under different attacks

As seen in Table 3, after applying attacks on watermarked EEG signals, it is observed that the values of BER are very low (less than 3%) while the values of NC is very high (close to one), which implies extracted watermark is very similar to the original watermark. Figure 7(b-e) shows the extracted watermark after attacks. Therefore, this indicates that the robustness of the proposed scheme is very good. In addition, BER < 3% could be corrected with the use of error correcting codes (Wicker 1995).

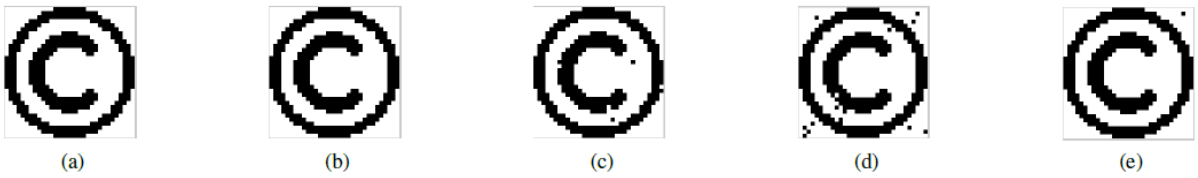

Figure 7. Result of watermark extraction at channel Fp1 of subject 01. (a) Original watermark. (b) Free attack. (c) Noise Addition. (d) Low-pass filtering. (e) Re-sampling.

6.3 Security

For a secure watermarking scheme for biomedical data, security is a very important issue. To improve confidentiality, the key space should be large enough to boost confidentiality and render attacks, especially brute force attacks, are impossible. Therefore, secret keys are adopted for security purposes, the proposed method utilizes chaotic encryption. The embedding and extraction processes in the proposed watermarking scheme depends on the secret key K, it is impossible to malicious attack to detect the watermark without this key. Supposed attacker can have the information of the watermarked signal, watermarking algorithm, and the encryption function. However, attacker cannot generate the watermark

without secret key. In addition, the proposed method possesses the high robustness which against attack is very important for a secured watermarking scheme.

6.4 Error Analysis

To decide whether there is a watermark, the original watermark W is compared with the extracted watermark W'. If the BER between W and W' is less than a user-defined threshold T, there is a watermark, otherwise, there is no watermark. Actually, the threshold T is determined by the probability of the detection error due to a false alarm or rejection detection (Bhat et al., 2010). To determine the watermark threshold T, the false alarm and rejection are usually taken into consideration. The performance of a watermarking system is generally characterized by two types of errors (Fan and Wang 2009), the false-positive error and false-negative error. The false-positive error is the probability that an un-watermarked biomedical signal declared as watermarked by the decoder, while false-negative error is the probability that a watermarked biomedical signal declared as un-watermarked by the decoder. The probability of false-positive error PFP and probability of false-negative error P_{NE} can be computed as:

$$P_{FP} = 2^{-m} \sum_{h=\lceil \rho m \rceil}^{m} \binom{m}{h} \quad (16)$$

$$P_{FN} = \sum_{h=0}^{\lceil \rho m \rceil - 1} \left[\binom{h}{m} P^h (1-P)^{m-h} \right] \quad (17)$$

where $\binom{m}{h}$ is the binomial coefficient, m is the total number of watermark bits, h is the total number of matching bits, and P is probability of the difference between extracted watermark and original watermark ($w \neq w'$). According to (Bhat et al. 2010), the desired false alarm error must be smaller than 10-6 order of magnitude.

Figure 8 and Figure 9 show the probability of false-positive error P_{FP} and the probability of false-negative error P_{FN}, corresponding for watermark length m ∈ [0,100], which indicates that P_{FP} and P_{FN} approach 0 when watermark length m is larger than 20. In our method, m = 1024, hence the false positive probability and false negative probability are close to 0. Indeed, we have h = [(1-BER)×m], therefore BER less than 20% meets this demand. If we set BER = 20%, then ρ = 0.8. In our method, m = 1024, Eq. (16) gives P_{FP} = 2.6209 x 10-88, hence the false positive is close to zero.

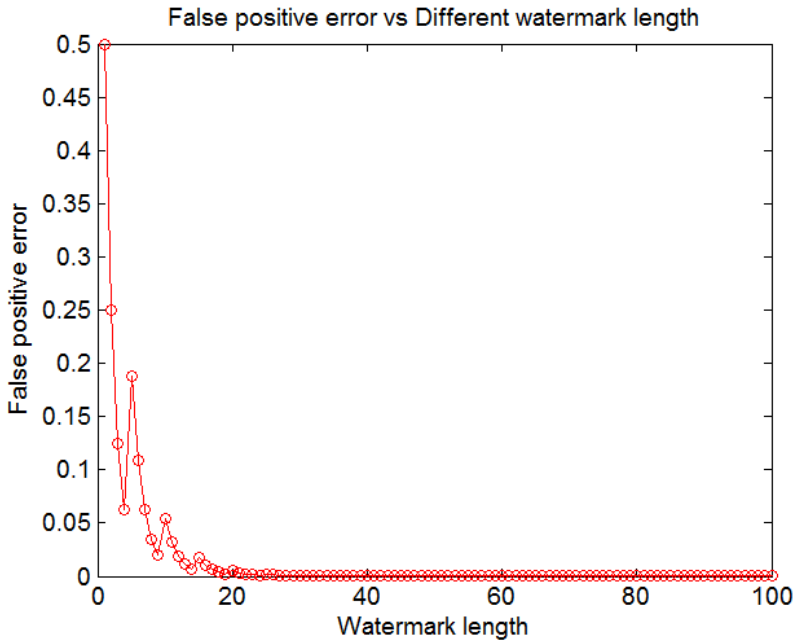

Figure 8. False Positive Error under different watermark length

In Eq. (17), the approximate value of P can be obtained from the BER under different attacks. As seen from Table 1 and Table 3, the average of BER is less than 3%, so P can be taken as 0.97. By substituting the values of m, p and P, Eq. (18) gives $P_{FN} = 1.5286 \times 10^{-102}$.

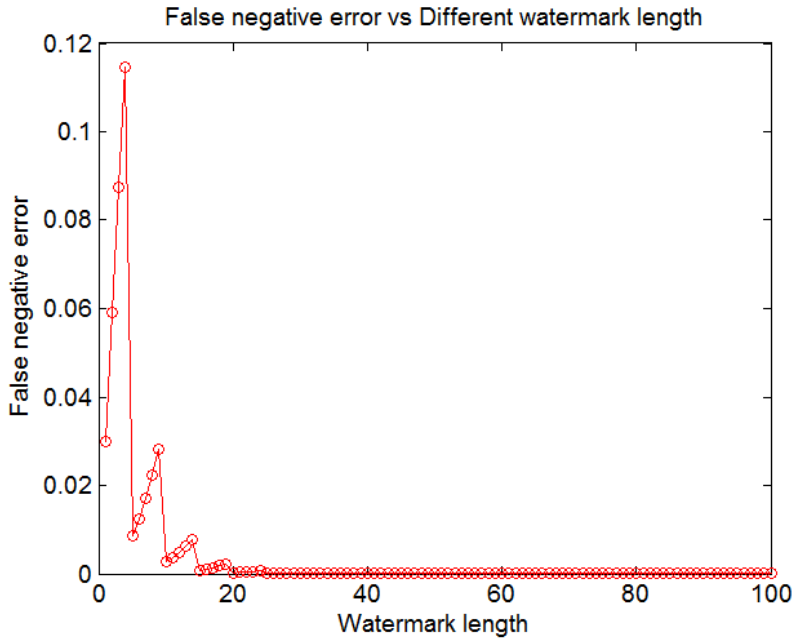

Figure 9. False Negative Error rate under different watermark length

In summary, our experimental results show that the proposed blind watermarking scheme for biomedical data has good imperceptibility and strong robustness against several different attacks such as noise addition, low-pass filtering, and re-sampling.

7 Conclusion and Future Work

An ownership protection based on the proposed optimized watermarking scheme for biomedical data in data mining has been developed. The proposed watermarking scheme is based on DWT with the exploration of SVD properties and DM quantization which make the scheme very robust to various common signal processing attacks. In our optimization process, PSO was used to make an optimal trade-off between imperceptibility and robustness through effective selection of quantization steps to locate the best parameters to insert the watermarks. The quantization steps were optimally adapted to achieve the most suitable performance for various biomedical data with different characteristics for medical application. In addition, our watermark scheme possesses the characteristic of blind extraction which does not require the original biomedical signal in extraction, thus reducing a lot of space for storing the original biomedical data and watermark. The experimental results have revealed that the proposed watermarking scheme achieves good imperceptibility and strong robustness against common signal processing. Our research can be used in e-healthcare application which need to share and transmit the biomedical data via network with ownership protection purpose. Since information can be extracted exactly, health information such as patient's data can be embedded in biomedical signal, reducing the consequences of health information thefts, increasing the data security, and saving storage space and bandwidth requirement for transmission of biomedical data. It is obvious that our study is also preferable to facilitate data management in health information management systems. In the future work, we will consider the following problems:

1. Enhancing privacy of biomedical data in data mining by tracing the source of an unauthorized release of biomedical data in networks. This is useful to monitor or trace back illegally produced copies of the data that may circulate. Tracing the source will

provide the information on who/where/how the biomedical data were accessed illegally.

2. Implementing the error coding code in watermark extraction, reducing the error and enhancing the performance of watermarking scheme.
3. Finding the most effective and suitable algorithm for optimal watermarking scheme. This work will save time and resources. It is necessary to investigate other evolutionary algorithms to enhance the performance with respect to the existing algorithms.
4. The proposed watermarking scheme for biomedical data is offline, it should be improved to carry out online scenarios.

References

Alhaqbani, B., and Fidge, C. 2008. "Privacy-Preserving Electronic Health Record Linkage Using Pseudonym Identifiers," *e-health Networking, Applications and Services, 2008. HealthCom 2008. 10th International Conference on*, pp. 108-117.

Arsalan, M., Malik, S. A., and Khan, A. 2012. "Intelligent Reversible Watermarking in Integer Wavelet Domain for Medical Images," *Journal of Systems and Software* (85:4), pp. 883-894.

Australia. Law Reform, C. 2002. *Principled Regulation: Report: Federal Civil & Administrative Penalties in Australia*. Australian Law Reform Commission.

Bender, W., Gruhl, D., Morimoto, N., and Lu, A. 1996. "Techniques for Data Hiding," *IBM systems journal* (35:3.4), pp. 313-336.

Bertino, E., Ooi, B. C., Yang, Y., and Deng, R. H. 2005. "Privacy and Ownership Preserving of Outsourced Medical Data," *21st International Conference on Data Engineering (ICDE'05)*, pp. 521-532.

Bhat, V., Sengupta, I., and Das, A. 2010. "An Adaptive Audio Watermarking Based on the Singular Value Decomposition in the Wavelet Domain," *Digital Signal Processing* (20:6), pp. 1547-1558.

Bhatnagar, G., and Wu, Q. M. J. 2013. "Biometrics Inspired Watermarking Based on a Fractional Dual Tree Complex Wavelet Transform," *Future Generation Computer Systems* (29:1), pp. 182-195.

Centers for, M., and Medicaid, S. 1996. "The Health Insurance Portability and Accountability Act of 1996 (Hipaa)," *Online at http://www.cms.hhs.gov/hipaa*.

Chang, C.-C., Tsai, P., and Lin, C.-C. 2005. "Svd-Based Digital Image Watermarking Scheme," *Pattern Recognition Letters* (26:10), pp. 1577-1586.

Chen, B., and Wornell, G. W. 2001. "Quantization Index Modulation Methods for Digital Watermarking and Information Embedding of Multimedia," *Journal of VLSI Signal Processing Systems* (27:1-2), pp. 7-33.

Chen, K., and Ramabadran, T. V. 1994. "Near-Lossless Compression of Medical Images through Entropy-Coded Dpcm," *IEEE Transactions on Medical Imaging* (13:3), pp. 538-548.

Chen, S.-T., Guo, Y.-J., Huang, H.-N., Kung, W.-M., Tseng, K.-K., and Tu, S.-Y. 2014. "Hiding Patients Confidential Datainthe Ecg Signal Viaa Transform-Domain Quantization Scheme," *Journal of medical systems* (38:6), pp. 54-74.

Chen, T.-S., Chang, C.-C., and Hwang, M.-S. 1998. "A Virtual Image Cryptosystem Based Upon Vector Quantization," *Image Processing, IEEE Transactions on* (7:10), pp. 1485-1488.

Chen, Y., and Xu, H. 2013. "Privacy Management in Dynamic Groups: Understanding Information Privacy in Medical Practices," ACM, pp. 541-552.

Clearinghouse, P. R. 2017. "Chronology of Data Breaches: Security Breaches 2005–Present," *2005, https://www.privacyrights.org/data-breach*.

Cox, I., Miller, M., Bloom, J., Fridrich, J., and Kalker, T. 2007. *Digital Watermarking and Steganography*. Morgan Kaufmann.

Dorigo, M., Di Caro, G., and Gambardella, L. M. 1999. "Ant Algorithms for Discrete Optimization," *Artificial life* (5:2), pp. 137-172.

Engin, M., Cidam, O., and Engin, E. Z. 2005. "Wavelet Transformation Based Watermarking Technique for Human Electrocardiogram (Ecg)," *Journal of Medical Systems* (29:6), pp. 589-594.

Fakhari, P., Vahedi, E., and Lucas, C. 2011. "Protecting Patient Privacy from Unauthorized Release of Medical Images Using a Bio-Inspired Wavelet-Based Watermarking Approach," *Digital Signal Processing* (21:3), pp. 433-446.

Fan, M., and Wang, H. 2009. "Chaos-Based Discrete Fractional Sine Transform Domain Audio Watermarking Scheme," *Computers and Electrical Engineering* (35:3), pp. 506-516.

Gupta, A. K., and Raval, M. S. 2012. "A Robust and Secure Watermarking Scheme Based on Singular Values Replacement," *Sadhana* (37:4), pp. 425-440.

Hänsch, K., and Serna, L. A. "Directive 95/46/Ec of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, Nov. 1995."

Hassan, R., Cohanim, B., De Weck, O., and Venter, G. 2005. "A Comparison of Particle Swarm Optimization and the Genetic Algorithm," *Proceedings of the 1st AIAA multidisciplinary design optimization specialist conference*, pp. 18-21.

He, X., Tseng, K.-K., Huang, H.-N., Chen, S.-T., Tu, S.-Y., Zeng, F., and Pan, J.-S. 2012. "Wavelet-Based Quantization Watermarking for Ecg Signals," *Computing, Measurement, Control and Sensor Network (CMCSN), 2012 International Conference on*, pp. 233-236.

Health, U. S. D. o., and Human, S. 2009. "Code of Federal Regulations. Title 45," *Public Welfare CFR* (46).

Heylen, K., and Dams, T. 2008. "An Image Watermarking Tutorial Tool Using Matlab," pp. 70750D-70751.

Holland, J. H. 1992. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT press.

Hu, F., Jiang, M., Wagner, M., and Dong, D.-C. 2007. "Privacy-Preserving Telecardiology Sensor Networks: Toward a Low-Cost Portable Wireless Hardware/Software Codesign," *IEEE Transactions on Information Technology in Biomedicine* (11:6), pp. 619-627.

Huang, C.-H., and Wu, J.-L. 2009. "Fidelity-Guaranteed Robustness Enhancement of Blind-Detection Watermarking Schemes," *Information Sciences* (179:6), pp. 791-808.

Ibaida, A., Khalil, I., and Van Schyndel, R. 2011. "A Low Complexity High Capacity Ecg Signal Watermark for Wearable Sensor-Net Health Monitoring System," *2011 Computing in Cardiology*, pp. 393-396.

Jahankhani, P., Kodogiannis, V., and Revett, K. 2006. "Eeg Signal Classification Using Wavelet Feature Extraction and Neural Networks," *Modern Computing, 2006. JVA'06. IEEE John Vincent Atanasoff 2006 International Symposium on*, pp. 120-124.

Jero, S. E., Ramu, P., and Ramakrishnan, S. 2014. "Discrete Wavelet Transform and Singular Value Decomposition Based Ecg Steganography for Secured Patient Information Transmission," *Journal of medical systems* (38:10), pp. 1-11.

Kamran, M., and Farooq, M. 2012. "An Information-Preserving Watermarking Scheme for Right Protection of Emr Systems," *IEEE Transactions on Knowledge and Data Engineering* (24:11), pp. 1950-1962.

Kaur, S., Singhal, R., Farooq, O., and Ahuja, B. S. 2010. "Digital Watermarking of Ecg Data for Secure Wireless Commuication," *Recent Trends in Information, Telecommunication and Computing (ITC), 2010 International Conference on*, pp. 140-144.

Kennedy, J. 2011. "Particle Swarm Optimization," in *Encyclopedia of Machine Learning*. Springer, pp. 760-766.

Kennedy, J. F., Kennedy, J., Eberhart, R. C., and Shi, Y. 2001. *Swarm Intelligence*. Morgan Kaufmann.

Ko, L.-T., Chen, J.-E., Shieh, Y.-S., Hsin, H.-C., and Sung, T.-Y. 2011. "Nested Quantization Index Modulation for Reversible Watermarking and Its Application to Healthcare Information Management Systems," *Computational and mathematical methods in medicine* (2012).

Koelstra, S., Mu, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. 2012. "Deap: A Database for Emotion Analysis; Using Physiological Signals," *Affective Computing, IEEE Transactions on* (3:1), pp. 18-31.

Kong, X., and Feng, R. 2001. "Watermarking Medical Signals for Telemedicine," *IEEE Transactions on Information Technology in Biomedicine* (5:3), pp. 195-201.

Kumsawat, P. 2010. "A Genetic Algorithm Optimization Technique for Multiwavelet-Based Digital Audio Watermarking," *EURASIP Journal on Advances in Signal Processing* (2010:1), p. 1.

Lai, C.-C. 2011. "A Digital Watermarking Scheme Based on Singular Value Decomposition and Tiny Genetic Algorithm," *Digital Signal Processing* (21:4), pp. 522-527.

Latifpour, H., Mosleh, M., and Kheyrandish, M. 2015. "An Intelligent Audio Watermarking Based on Knn Learning Algorithm," *International Journal of Speech Technology* (18:4), pp. 697-706.

Lee, W.-B., and Lee, C.-D. 2008. "A Cryptographic Key Management Solution for Hipaa Privacy/Security Regulations," *IEEE Transactions on Information Technology in Biomedicine* (12:1), pp. 34-41.

Lei, B., Song, I., and Rahman, S. A. 2013. "Robust and Secure Watermarking Scheme for Breath Sound," *Journal of Systems and Software* (86:6), pp. 1638-1649.

Li, Z.-R., Chang, E.-C., Huang, K.-H., and Lai, F. 2011. "A Secure Electronic Medical Record Sharing Mechanism in the Cloud Computing Platform," pp. 98-103.

Loukhaoukha, K., Chouinard, J.-Y., and Taieb, M. H. 2011. "Optimal Image Watermarking Algorithm Based on Lwt-Svd Via Multi-Objective Ant Colony Optimization," *Journal of Information Hiding and Multimedia Signal Processing* (2-6:4), pp. 303-319.

Maglogiannis, I., Kazatzopoulos, L., Delakouridis, K., and Hadjiefthymiades, S. 2009. "Enabling Location Privacy and Medical Data Encryption in Patient Telemonitoring Systems," *IEEE Transactions on Information Technology in Biomedicine* (13:6), pp. 946-954.

Malasri, K., and Wang, L. 2007. "Addressing Security in Medical Sensor Networks," ACM, pp. 7-12.

Mishra, A., Agarwal, C., Sharma, A., and Bedi, P. 2014. "Optimized Gray-Scale Image Watermarking Using Dwt–Svd and Firefly Algorithm," *Expert Systems with Applications* (41:17), pp. 7858-7867.

Mousavi, S. M., Naghsh, A., and Abu-Bakar, S. A. R. 2014. "Watermarking Techniques Used in Medical Images: A Survey," *Journal of digital imaging* (27:6), pp. 714-729.

Orhan, U., Hekim, M., and Ozer, M. 2011. "Eeg Signals Classification Using the K-Means Clustering and a Multilayer Perceptron Neural Network Model," *Expert Systems with Applications* (38:10), pp. 13475-13481.

Percival, D. B., and Walden, A. T. 2006. *Wavelet Methods for Time Series Analysis*. Cambridge university press.

Pham, T. D., Tran, D., and Ma, W. 2015. "A Proposed Blind Dwt-Svd Watermarking Scheme for Eeg Data," *International Conference on Neural Information Processing*, pp. 69-76.

Planitz, B., and Maeder, A. 2005. "Medical Image Watermarking: A Study on Image Degradation," *Proc. Australian Pattern Recognition Society Workshop on Digital Image Computing, WDIC*.

Prior, F., Ingeholm, M. L., Levine, B. A., and Tarbox, L. 2009. "Potential Impact of Hitech Security Regulations on Medical Imaging," IEEE, pp. 2157-2160

Ramu, P., Swaminathan, R., and others. 2016. "Imperceptibility—Robustness Tradeoff Studies for Ecg Steganography Using Continuous Ant Colony Optimization," *Expert Systems with Applications* (49), pp. 123-135.

Run, R.-S., Horng, S.-J., Lai, J.-L., Kao, T.-W., and Chen, R.-J. 2012. "An Improved Svd-Based Watermarking Technique for Copyright Protection," *Expert Systems with applications* (39:1), pp. 673-689.

Ruotsalainen, P. 2010. "Privacy and Security in Teleradiology," *European Journal of Radiology* (73:1), pp. 31-35.

Smith, R., and Gotel, O. 2007. "Using a Game to Introduce Lightweight Requirements Engineering," IEEE, pp. 379-380.

Solutions, V. E. 2015. "Data Breach Investigations Report," *Verizon, Report*.

Sriyingyong, N., and Attakitmongcol, K. 2006. "Wavelet-Based Audio Watermarking Using Adaptive Tabu Search," *2006 1st International Symposium on Wireless Pervasive Computing*, pp. 1-5.

Storn, R., and Price, K. 1997. "Differential Evolution–a Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," *Journal of global optimization* (11:4), pp. 341-359.

Tefas, A., Nikolaidis, N., and Pitas, I. 2009. "Image Watermarking-Chapter 22: Techniques and Applications".

Trelea, I. C. 2003. "The Particle Swarm Optimization Algorithm: Convergence Analysis and Parameter Selection," *Information processing letters* (85:6), pp. 317-325.

Tsai, H.-H., Jhuang, Y.-J., and Lai, Y.-S. 2012. "An Svd-Based Image Watermarking in Wavelet Domain Using Svr and Pso," *Applied Soft Computing* (12:8), pp. 2442-2453.

Van Schyndel, R. G., Tirkel, A. Z., and Osborne, C. F. 1994. "A Digital Watermark," IEEE, pp. 86-90.

Voyatzis, G., and Pitas, I. 1999. "The Use of Watermarks in the Protection of Digital Multimedia Products," *Proceedings of the IEEE* (87:7), pp. 1197-1207.

Wang, H., Peng, D., Wang, W., Sharif, H., Chen, H.-H., and Khoynezhad, A. 2010. "Resource-Aware Secure Ecg Healthcare Monitoring through Body Sensor Networks," *IEEE Wireless Communications* (17:1), pp. 1536-1284.

Wang, J., Peng, H., and Shi, P. 2011. "An Optimal Image Watermarking Approach Based on a Multi-Objective Genetic Algorithm," *Information Sciences* (181:24), pp. 5501-5514.

Westin, K. 2015. "Encryption Would't Have Stopped Anthem's Data Breach." MIT Technology Review, Febraury.

Wicker, S. B. 1995. *Error Control Systems for Digital Communication and Storage*. Prentice hall Englewood Cliffs.

Zain, J. M., and Clarke, M. 2011. "Reversible Region of Non-Interest (Roni) Watermarking for Authentication of Dicom Images," *arXiv preprint arXiv:1101.1603*.

Zheng, D., Liu, Y., Zhao, J., and Saddik, A. E. 2007. "A Survey of Rst Invariant Image Watermarking Algorithms," *ACM Computing Surveys (CSUR)* (39:2), p. 5.